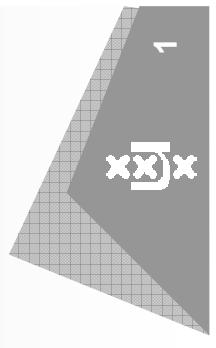


Facing the Challenge

A Divide and Conquer Approach to Concept Detection

C.G.M. Snoek, J.C. van Gemert, Th. Gevers, B. Huurnink, D.C. Koelma,
K.E.A. van de Sande, A.W.M. Smeulders, C.J. Veenman, M. Worring

Intelligent Systems Lab Amsterdam,
University of Amsterdam, The Netherlands



NIST TRECVID benchmark

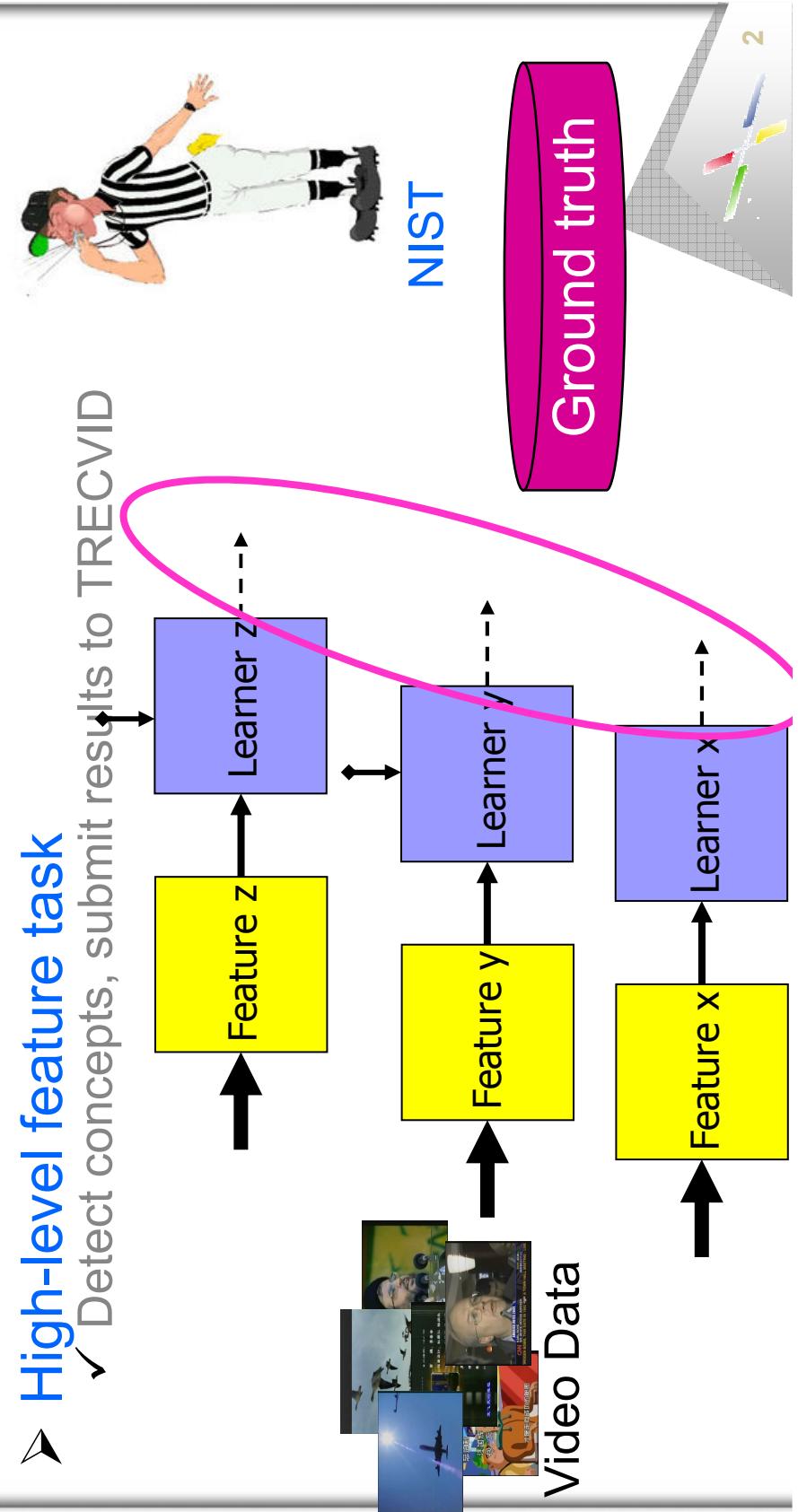
- Introduction
- Challenge
- Visual-only
- Results
- Lessons

► Benchmark objectives

- ✓ Promote progress in video retrieval research
- ✓ Provide common dataset (shots, annotations, key frames)
- ✓ Use open, metrics-based evaluation

► High-level feature task

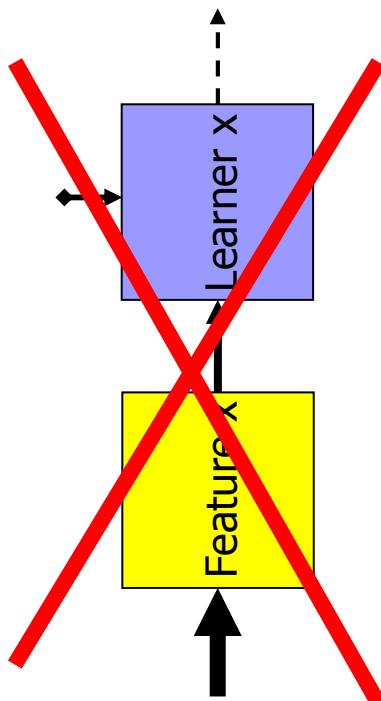
- ✓ Detect concepts, submit results to TRECVID



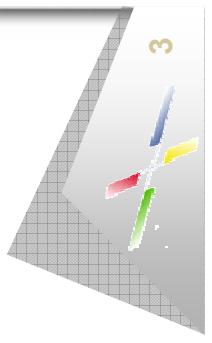
Benchmark limitations

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

- **Focus is on the final result**
 - ✓ TRECVID judges **relative** merit of indexing methods
 - ✓ Ignores repeatability of intermediate analysis steps
- **Systems are becoming more complex**
 - ✓ Typically combining several features and learning methods



- **Share video data, not the annotations and features**
 - ✓ Component-based optimization and comparison impossible

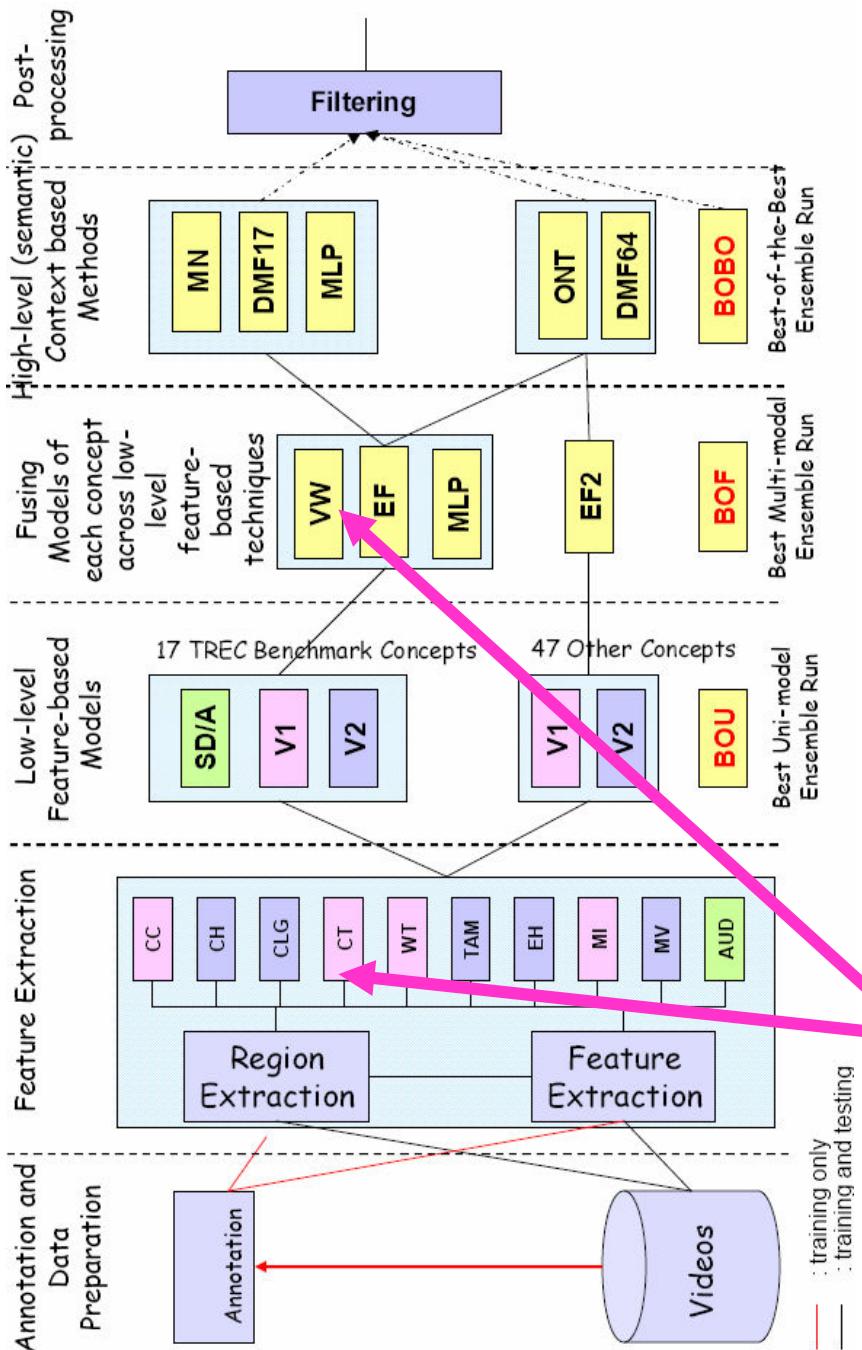


IBM

Semantic video indexing

Analysis Pipeline

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

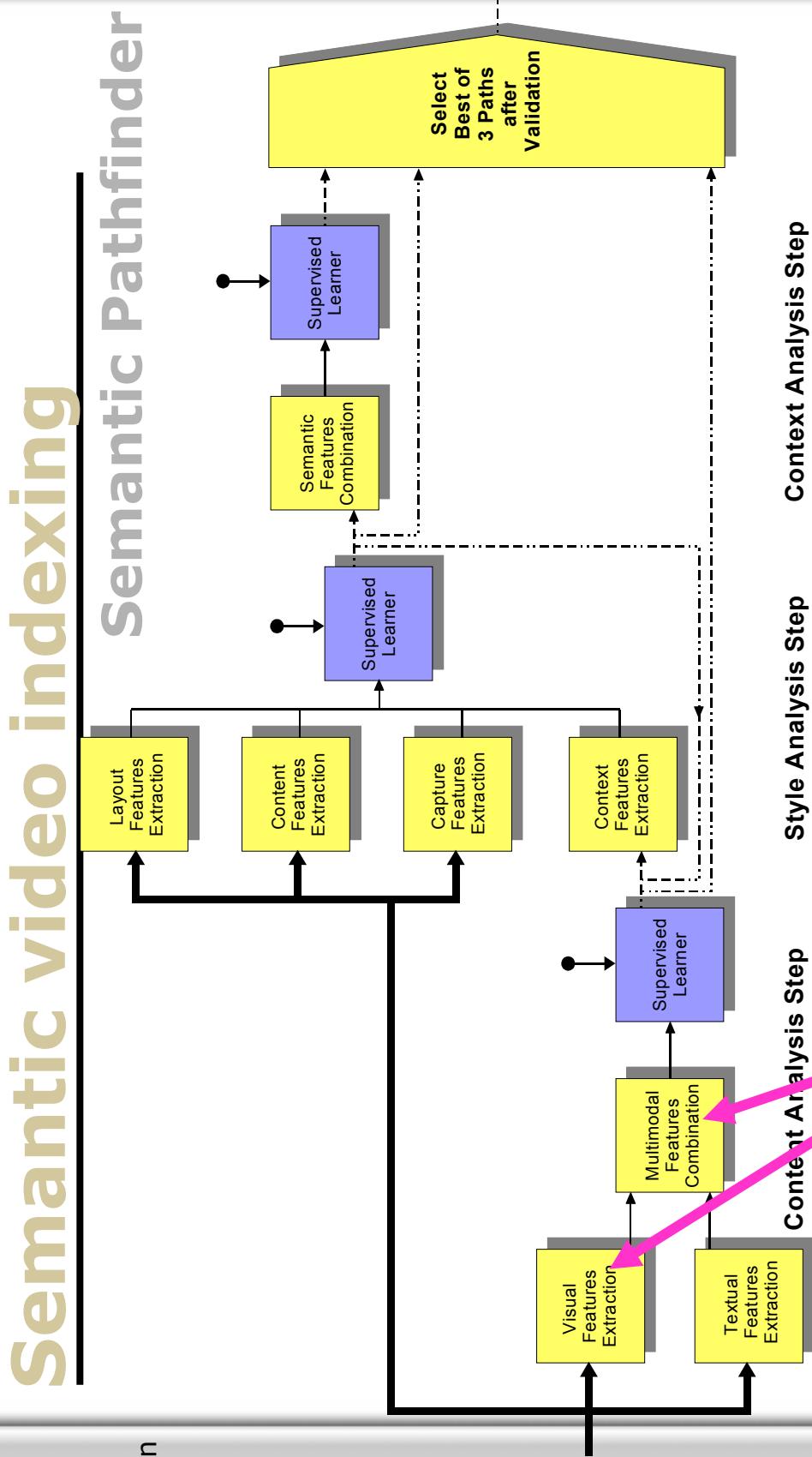


What is contribution of these components?

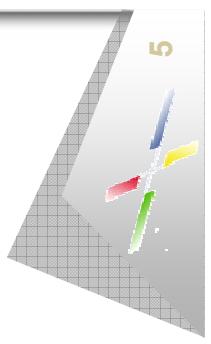
MediaMill

Semantic video indexing

- Introduction
- Challenge
- Visual-only
- Results
- Lessons



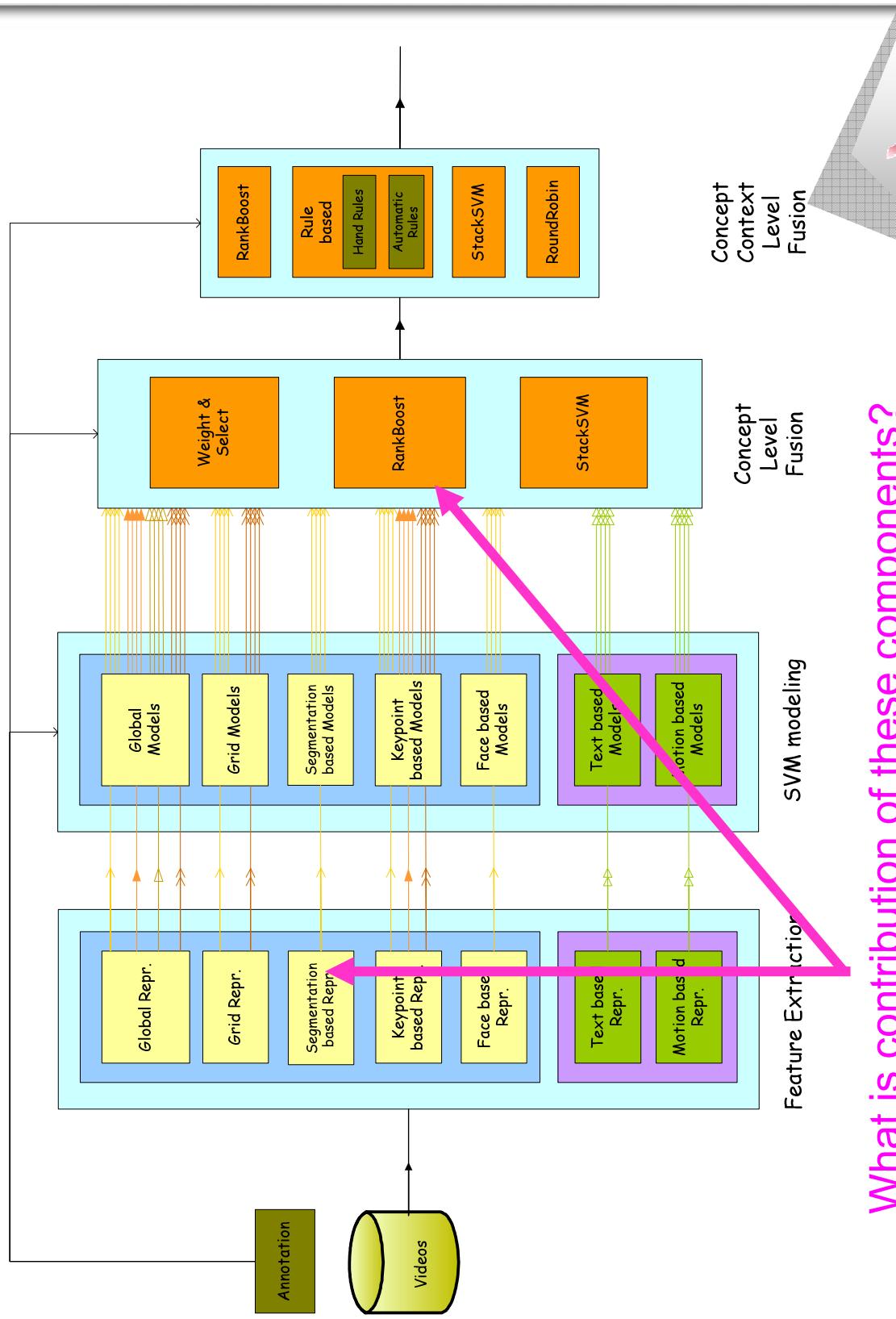
What is contribution of these components?



Tsinghua

Semantic video indexing

- Introduction
- Challenge
- Visual-only
- Results
- Lessons



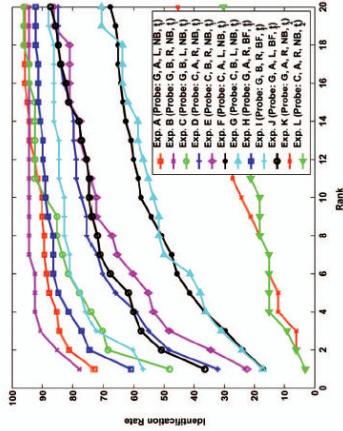
What is contribution of these components?

Lessons from computer vision

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

► Gait-based identification of humans

- ✓ Standardize common components



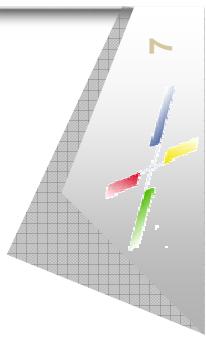
- ✓ What is needed?

- ❖ annotated data set
- ❖ baseline implementation
- ❖ baseline results

Challenge Problem

► We propose concept detection Challenge Problem

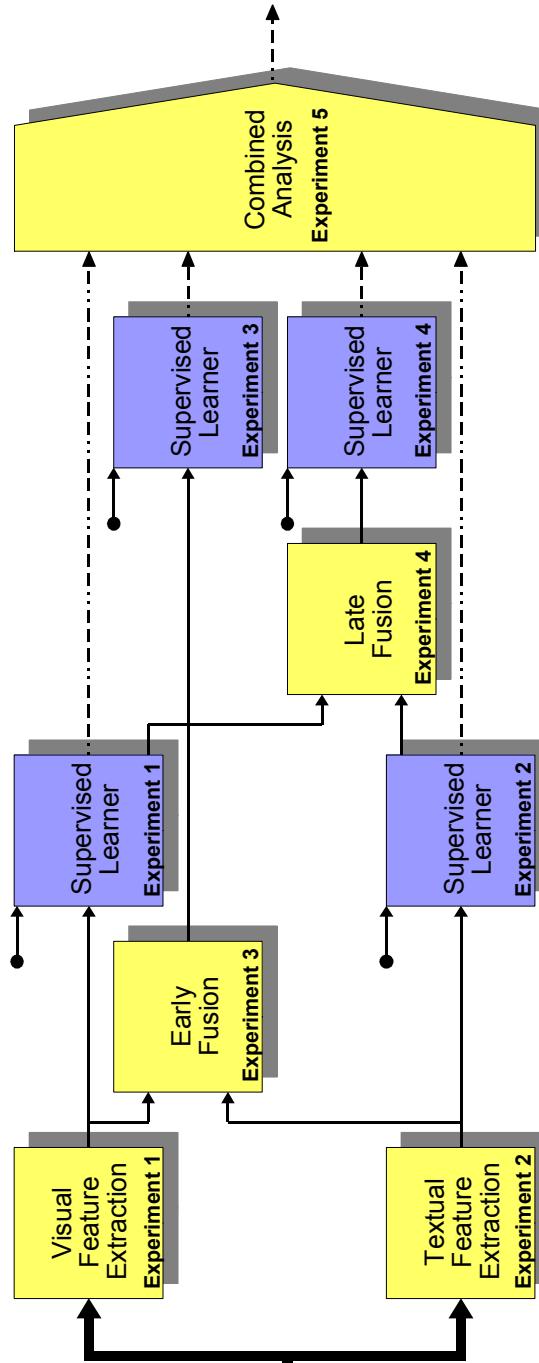
- ✓ Evaluates semantic video indexing methods
- ✓ Allows for component-based optimization
- ✓ Offers reference during feature development



The MediaMill Challenge

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

- Given a large video data set, we provide:
 - ✓ A lexicon of **101** annotated concepts
 - ✓ 5 pre-cooked multimedia analysis experiments
 - ✓ Baseline detector algorithms
 - ✓ Baseline performance



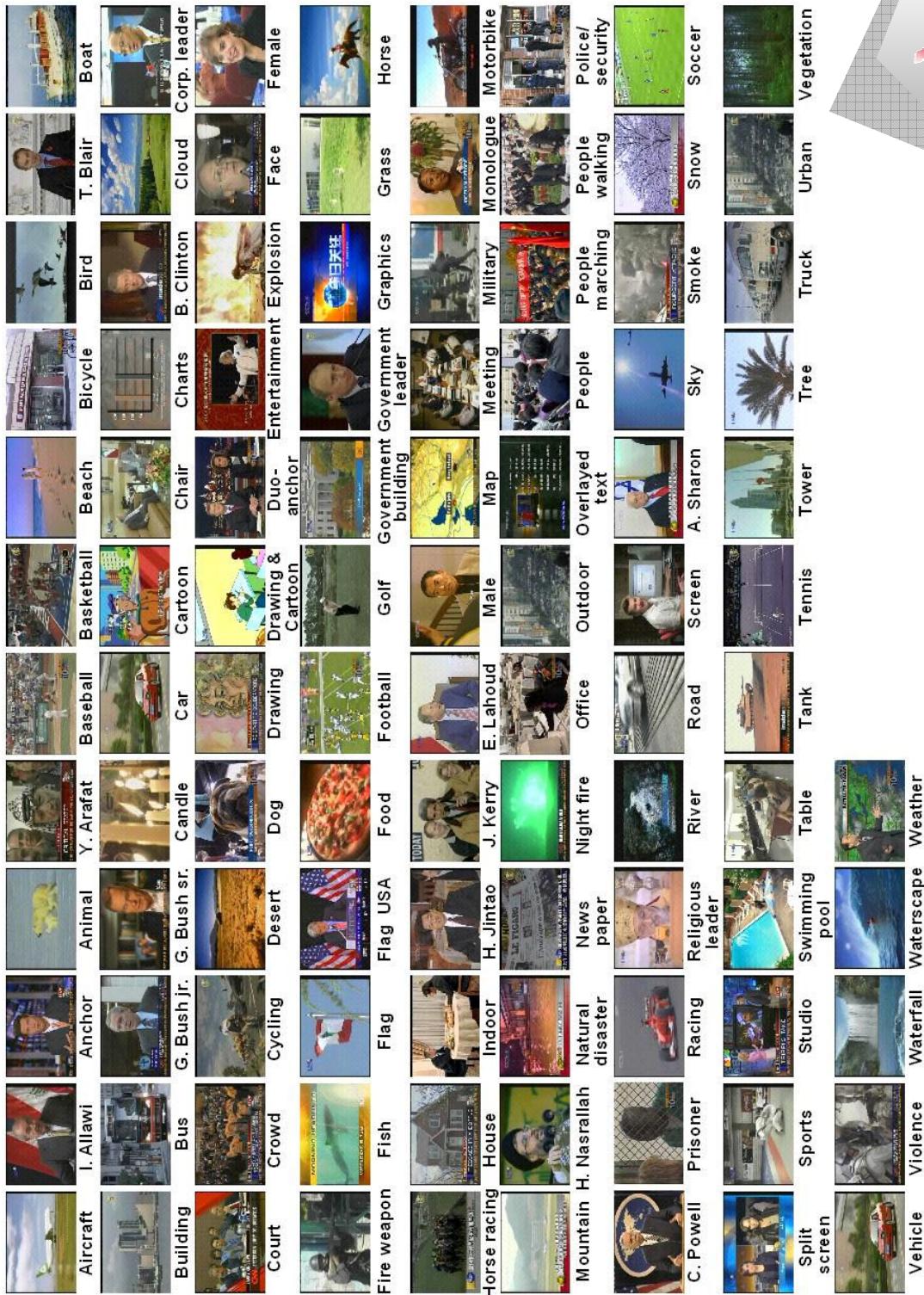
- Can the multimedia research community improve?



MediaMill

Annotated 101 concept lexicon

- Introduction
- Challenge
- Visual-only
- Results
- Lessons



Baseline results

■ Introduction

► Reference for comparison

■ Visual-only

■ Results

■ Lessons

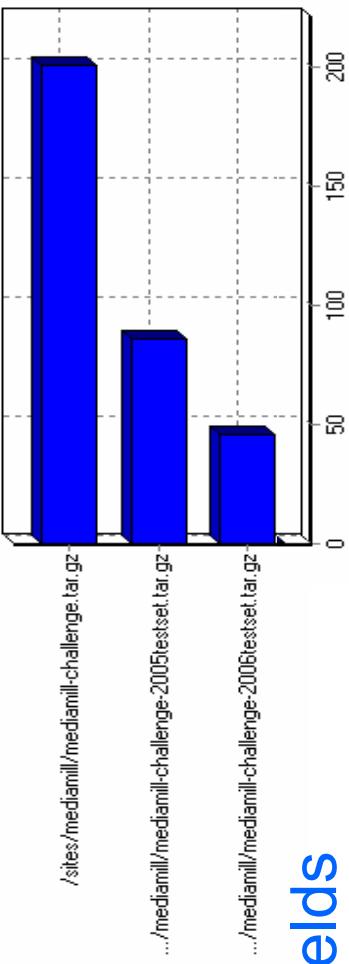
	Ground Truth	Challenge Experiments					Challenge Experiments	
		Train (%)	Test (%)	1	2	3	4	5
Concept	77.67	75.87	0.831	0.817	0.890	0.840	0.840	0.52
1 People	64.15	62.37	0.895	0.737	0.892	0.890	0.890	0.52
2 Face	36.33	34.30	0.669	0.533	0.642	0.666	0.669	0.53
3 Overlaid text								0.057
4 Outdoor	32.68	38.33	0.688	0.579	0.709	0.691	0.691	0.53
5 Entertainment	19.64	12.55	0.165	0.179	0.267	0.146	0.179	0.62
6 Indoor	19.59	21.20	0.593	0.480	0.592	0.606	0.592	0.62
7 Studio	13.66	14.20	0.636	0.490	0.654	0.651	0.654	0.62
8 People walking	13.61	16.83	0.363	0.294	0.338	0.296	0.338	0.68
9 Utan	11.78	8.80	0.222	0.178	0.195	0.201	0.195	0.44
10 Crowd	11.48	16.12	0.480	0.288	0.490	0.446	0.490	0.36
11 Sky	10.77	11.38	0.478	0.218	0.496	0.463	0.496	0.34
12 GovernmentLeader	9.35	7.87	0.213	0.213	0.222	0.236	0.213	0.33
13 Violence	8.07	9.75	0.317	0.301	0.334	0.237	0.334	0.22
14 Road	7.76	6.60	0.195	0.138	0.212	0.188	0.195	0.27
15 Vehicle	7.61	8.53	0.221	0.167	0.271	0.190	0.271	0.30
16 Building	6.86	11.16	0.316	0.154	0.233	0.291	0.154	0.15
17 Male	5.71	2.38	0.066	0.034	0.068	0.069	0.086	0.29
18 Anchor	5.09	4.85	0.631	0.201	0.620	0.618	0.631	0.27
19 Car	4.87	5.93	0.262	0.118	0.246	0.215	0.252	0.19
20 Meeting	4.53	4.25	0.257	0.158	0.211	0.257	0.257	0.12
21 Female	4.38	2.11	0.086	0.020	0.061	0.068	0.086	0.26
22 Military	4.14	6.58	0.217	0.205	0.235	0.203	0.235	0.31
23 Vegetation	3.87	4.64	0.183	0.051	0.161	0.183	0.183	0.21
24 Sports	3.76	2.61	0.304	0.267	0.231	0.308	0.304	0.20
25 Monologue	3.10	2.33	0.094	0.051	0.074	0.081	0.094	0.23
26 Graphics	2.89	3.48	0.365	0.275	0.379	0.367	0.365	0.20
27 Computer leader	2.57	1.30	0.016	0.020	0.014	0.018	0.020	0.19
28 Watercolor	2.31	1.89	0.150	0.079	0.134	0.142	0.134	0.18
29 People marching	1.93	4.13	0.228	0.087	0.267	0.199	0.267	0.17
30 Soccer	1.67	0.29	0.503	0.000	0.079	0.372	0.503	0.16
31 Mountain	1.64	1.01	0.141	0.022	0.092	0.157	0.141	0.14
32 G. Bush [t]	1.61	0.54	0.062	0.065	0.040	0.060	0.062	0.05
33 Office	1.56	1.75	0.077	0.024	0.045	0.037	0.024	0.12
34 Screen	1.53	1.90	0.101	0.063	0.058	0.121	0.063	0.10
35 Flag	1.26	1.12	0.189	0.029	0.120	0.166	0.189	0.09
36 Track	1.16	1.02	0.038	0.019	0.042	0.038	0.019	0.10
37 Map	1.16	1.21	0.220	0.213	0.131	0.407	0.476	0.08
38 Smoke	1.13	2.14	0.250	0.103	0.366	0.149	0.260	0.08
39 Animal	1.00	0.91	0.209	0.204	0.199	0.239	0.199	0.17
40 Weather	0.99	1.25	0.405	0.270	0.701	0.566	0.730	0.08
41 Aircraft	0.99	0.94	0.073	0.033	0.115	0.030	0.115	0.08
42 Police/Security	0.92	0.77	0.012	0.053	0.082	0.017	0.053	0.07
43 Flag USA	0.92	0.94	0.227	0.035	0.157	0.184	0.227	0.05
44 Grass	0.90	0.59	0.084	0.004	0.028	0.054	0.064	0.06
45 Cloud	0.87	1.54	0.117	0.042	0.078	0.129	0.117	0.05
46 Split screen	0.86	0.60	0.630	0.160	0.321	0.566	0.630	0.19
47 Desert	0.81	1.44	0.103	0.032	0.093	0.052	0.032	0.05
48 Natural disaster	0.81	0.93	0.055	0.091	0.139	0.084	0.091	0.04
49 Boat	0.78	0.54	0.096	0.109	0.083	0.020	0.109	0.03
50 Tree	0.78	0.84	0.124	0.011	0.063	0.087	0.124	0.01
51 Charts	0.76	0.61	0.327	0.301	0.254	0.355	0.327	0.01

Using the Challenge

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

► **Challenge donation to TRECVID community**

- ✓ 80+ participants downloaded the data
- ✓ Some even used it for their TRECVID submission

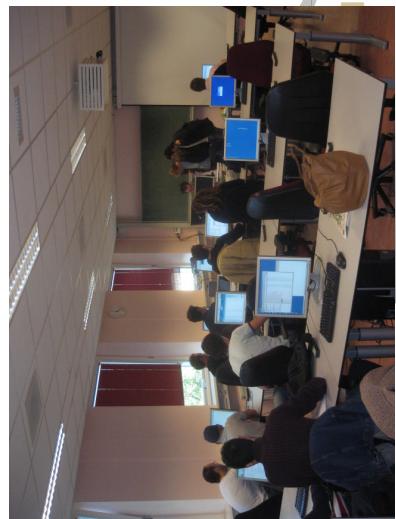


► **Used in other fields**

- ✓ Concept detectors used at INEX benchmark
- ✓ Referenced as standard dataset on LIBSVM website

► **Usage in education**

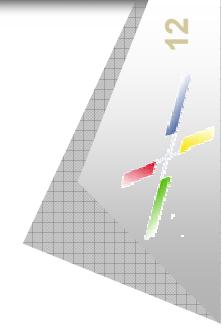
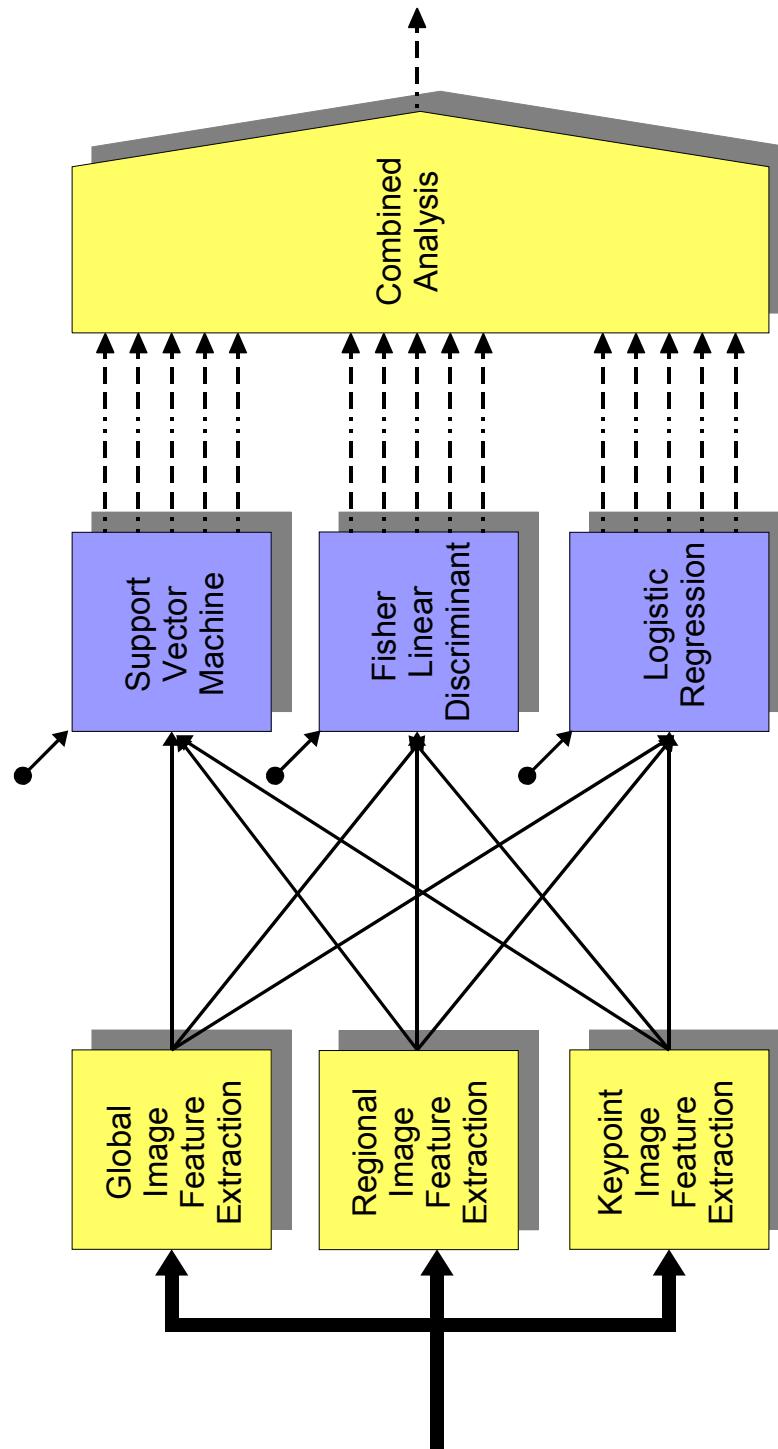
- ✓ M.Sc. Student lab course
- ✓ Lowers threshold to enter field



MediaMill TRECVID 2006 approach

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

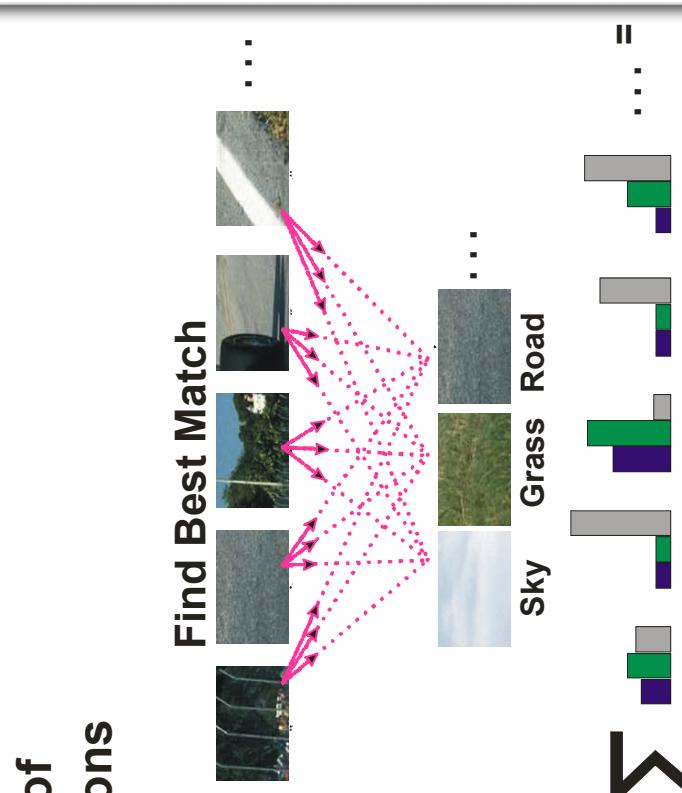
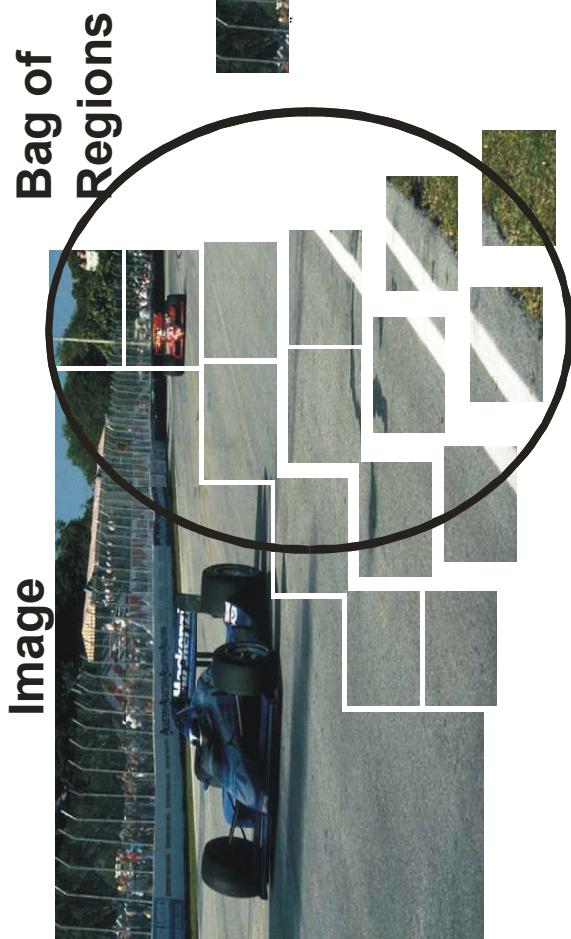
- Emphasize visual-only analysis
- ✓ Viz.: experiment 1 of the MediaMill Challenge



References:
Jurie & Triggs, ICCV 2005, Fei-Fei et al. CVPR 2005,
Sudderth et al. NIPS 2005, ...

Decompose image into proto-concepts

- Introduction
- Challenge
- Visual-only
- Results
- Lessons



Proto-concept Similarity Distribution:

► Problem

- ✓ Assumes one proto-concept per region is sufficient

► Solution

- ✓ Use similarity distribution

Global image features

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

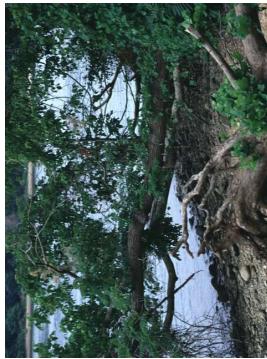
► Color Invariant features

- ✓ Robust to Shadows

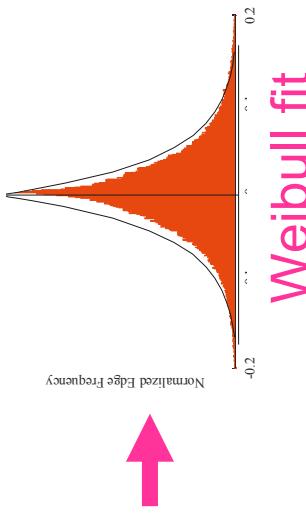
► Histogram of Edges

■ Lessons

- ✓ Resembles Weibull distribution
- Original



■ X-edges histogram



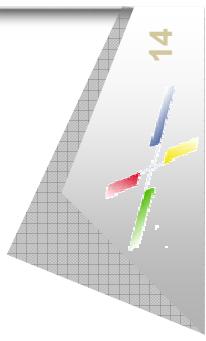
► Weibull depends on two parameters only

- ✓ Substantial data reduction in feature vector

► Proto-concept similarity measure

- ✓ Based on Weibull parameters

References:
Geusebroek et al, PAMI 2001
Geusebroek et al, IJCV 2005



Regional image features

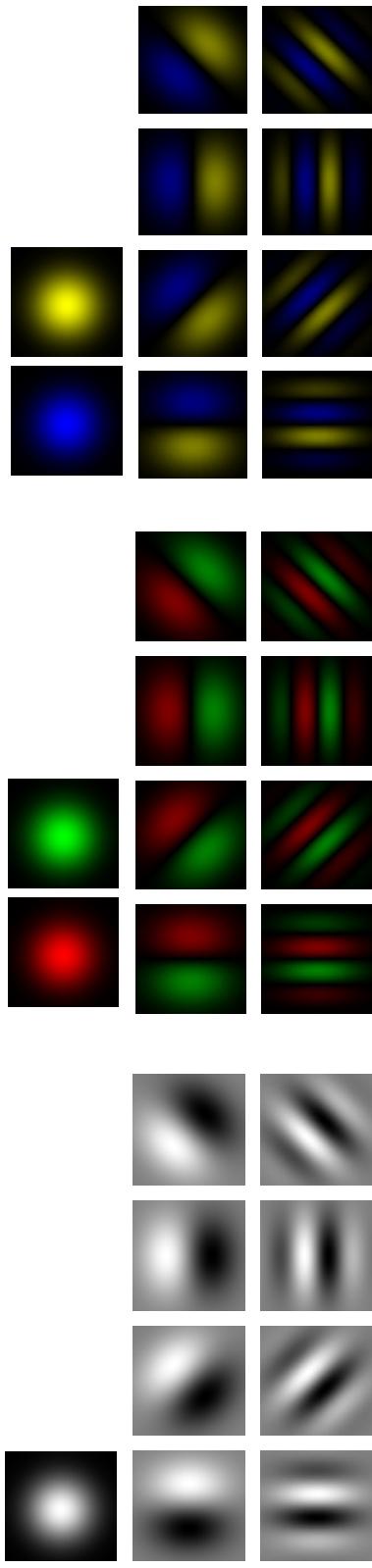
- Introduction
- Challenge
- Visual-only
- Results
- Lessons

➤ We also use Weibull fit on region level



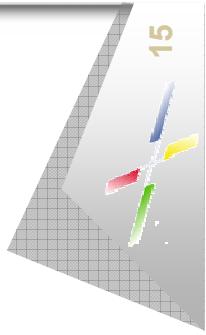
➤ Histogram of Gabor filter responses

- ✓ zero order = color histogram
- ✓ higher order = texture



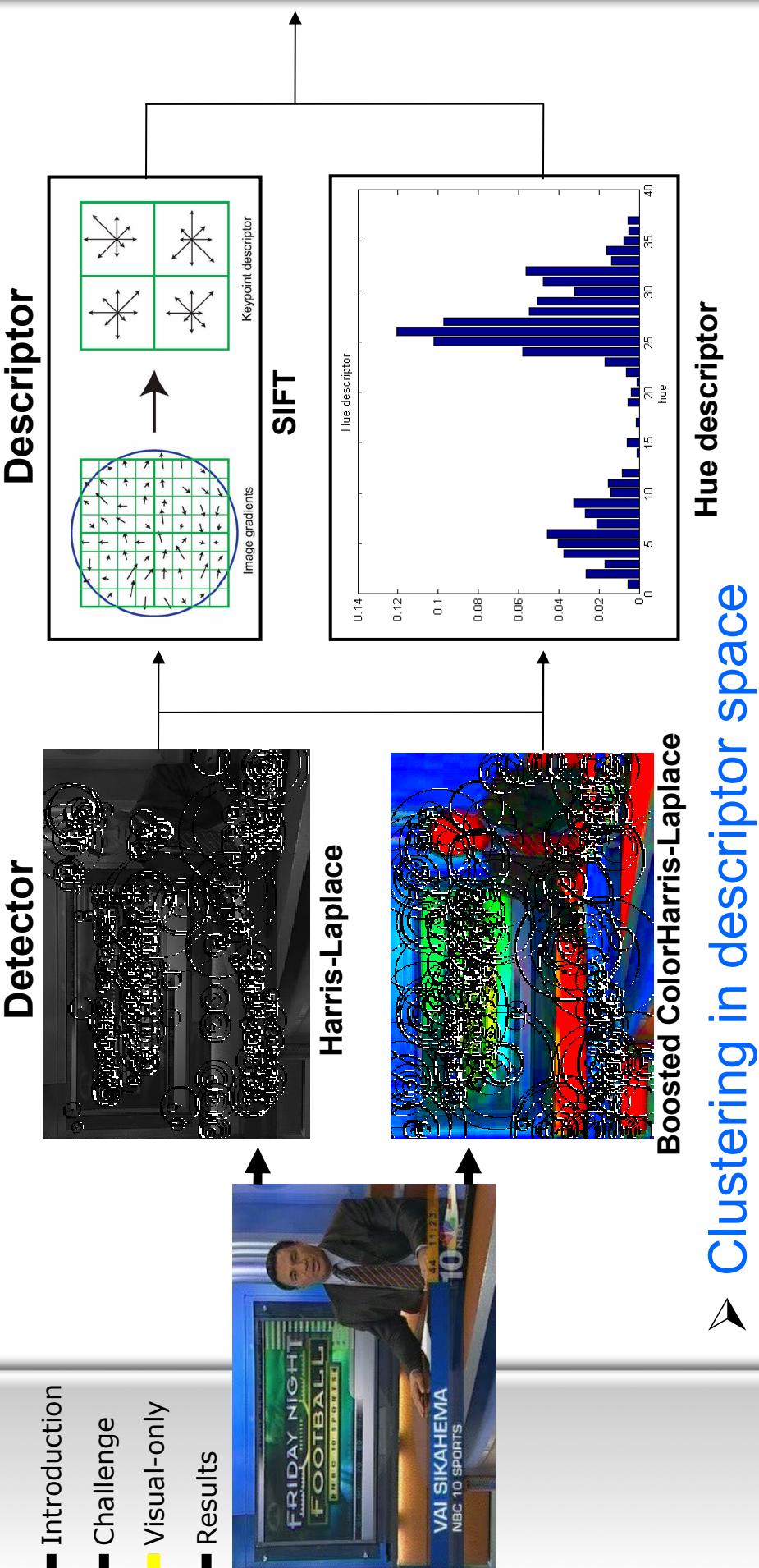
➤ Proto-concept similarity measure

- ✓ Histogram intersection



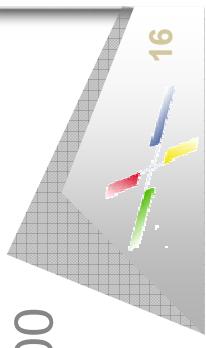
Keypoint image features

- Introduction
- Challenge
- Visual-only
- Results



► Clustering in descriptor space

- ✓ yields ‘proto-concept’ codebook
- ✓ 10 codebook elements per concept, totalling ~400
- ✓ dissimilarity codebook/descriptors yields vector

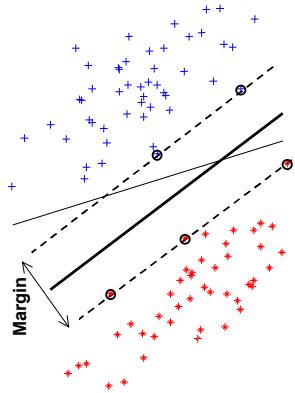


Supervised learners

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

► Support vector machine

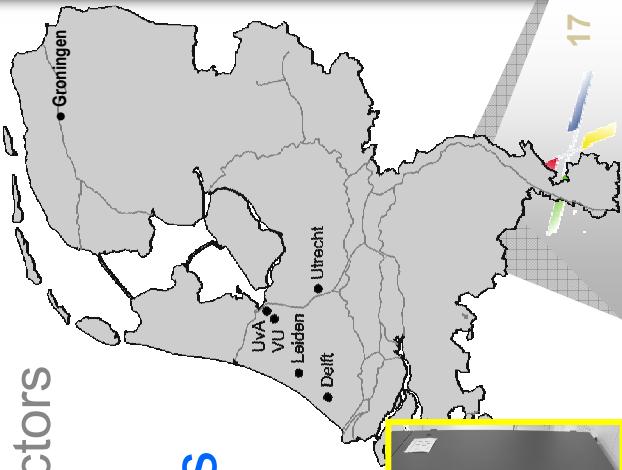
- ✓ Maximizes margin between two classes
- ✓ Problematic when data not balanced
- ✓ Solution: expensive parameter tuning
 - ❖ We experiment with several settings
 - ❖ We use 3 supercomputer clusters



$$p(\omega|\vec{s}) = \frac{1}{1 + \exp(\alpha\gamma(\vec{s}) + \beta)}$$

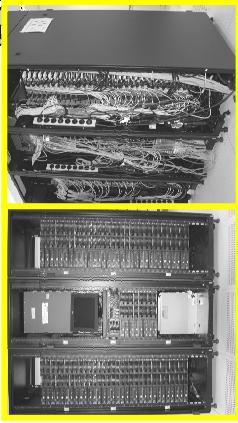
► Problem: SVMs are accurate but demanding

- ✓ Convergence too slow for large feature vectors



► Solution: try less demanding classifiers

- ✓ Logistic regression
- ✓ Fisher linear discriminant



visual-only analysis results

on Challenge

- Introduction
- Challenge
- Visual-only
 - ✓ Extensive and repeatable experiments by
 - ✓ modular distribution of our efforts
 - ✓ emphasizing features, classifiers, and settings
- Results
- Lessons

✓ Overview in notebook

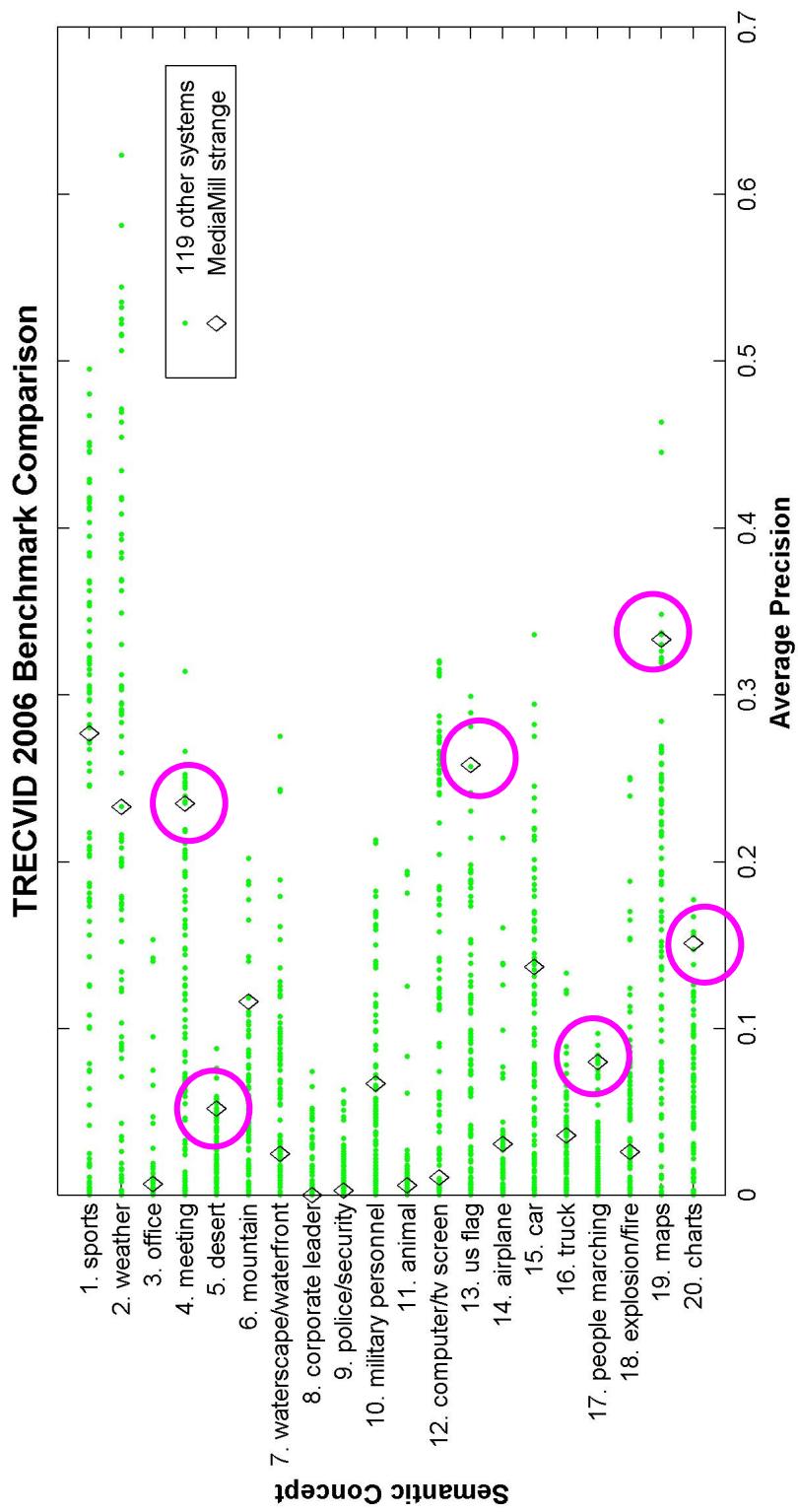
‘Best-of’ conquers the baseline by more than 50% MAP

Concept	Support Vector Machine						Logistic Regression						Fisher Linear Discriminant																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
W₀	W₁	W₂	W₃	W₄	W₅	W₆	W₇	W₈	W₉	W₁₀	W₁₁	W₁₂	W₁₃	W₁₄	W₁₅	W₁₆	W₁₇	W₁₈	W₁₉	W₂₀	W₂₁	W₂₂	W₂₃	W₂₄	W₂₅	W₂₆	W₂₇	W₂₈	W₂₉	W₃₀	W₃₁	W₃₂	W₃₃	W₃₄	W₃₅	W₃₆	W₃₇	W₃₈	W₃₉	W₄₀	W₄₁	W₄₂	W₄₃	W₄₄	W₄₅	W₄₆	W₄₇	W₄₈	W₄₉	W₅₀	W₅₁	W₅₂	W₅₃	W₅₄	W₅₅	W₅₆	W₅₇	W₅₈	W₅₉	W₆₀	W₆₁	W₆₂	W₆₃	W₆₄	W₆₅	W₆₆	W₆₇	W₆₈	W₆₉	W₇₀	W₇₁	W₇₂	W₇₃	W₇₄	W₇₅	W₇₆	W₇₇	W₇₈	W₇₉	W₈₀	W₈₁	W₈₂	W₈₃	W₈₄	W₈₅	W₈₆	W₈₇	W₈₈	W₈₉	W₉₀	W₉₁	W₉₂	W₉₃	W₉₄	W₉₅	W₉₆	W₉₇	W₉₈	W₉₉	W₁₀₀	W₁₀₁	W₁₀₂	W₁₀₃	W₁₀₄	W₁₀₅	W₁₀₆	W₁₀₇	W₁₀₈	W₁₀₉	W₁₁₀	W₁₁₁	W₁₁₂	W₁₁₃	W₁₁₄	W₁₁₅	W₁₁₆	W₁₁₇	W₁₁₈	W₁₁₉	W₁₂₀	W₁₂₁	W₁₂₂	W₁₂₃	W₁₂₄	W₁₂₅	W₁₂₆	W₁₂₇	W₁₂₈	W₁₂₉	W₁₃₀	W₁₃₁	W₁₃₂	W₁₃₃	W₁₃₄	W₁₃₅	W₁₃₆	W₁₃₇	W₁₃₈	W₁₃₉	W₁₄₀	W₁₄₁	W₁₄₂	W₁₄₃	W₁₄₄	W₁₄₅	W₁₄₆	W₁₄₇	W₁₄₈	W₁₄₉	W₁₅₀	W₁₅₁	W₁₅₂	W₁₅₃	W₁₅₄	W₁₅₅	W₁₅₆	W₁₅₇	W₁₅₈	W₁₅₉	W₁₆₀	W₁₆₁	W₁₆₂	W₁₆₃	W₁₆₄	W₁₆₅	W₁₆₆	W₁₆₇	W₁₆₈	W₁₆₉	W₁₇₀	W₁₇₁	W₁₇₂	W₁₇₃	W₁₇₄	W₁₇₅	W₁₇₆	W₁₇₇	W₁₇₈	W₁₇₉	W₁₈₀	W₁₈₁	W₁₈₂	W₁₈₃	W₁₈₄	W₁₈₅	W₁₈₆	W₁₈₇	W₁₈₈	W₁₈₉	W₁₉₀	W₁₉₁	W₁₉₂	W₁₉₃	W₁₉₄	W₁₉₅	W₁₉₆	W₁₉₇	W₁₉₈	W₁₉₉	W₂₀₀	W₂₀₁	W₂₀₂	W₂₀₃	W₂₀₄	W₂₀₅	W₂₀₆	W₂₀₇	W₂₀₈	W₂₀₉	W₂₁₀	W₂₁₁	W₂₁₂	W₂₁₃	W₂₁₄	W₂₁₅	W₂₁₆	W₂₁₇	W₂₁₈	W₂₁₉	W₂₂₀	W₂₂₁	W₂₂₂	W₂₂₃	W₂₂₄	W₂₂₅	W₂₂₆	W₂₂₇	W₂₂₈	W₂₂₉	W₂₃₀	W₂₃₁	W₂₃₂	W₂₃₃	W₂₃₄	W₂₃₅	W₂₃₆	W₂₃₇	W₂₃₈	W₂₃₉	W₂₄₀	W₂₄₁	W₂₄₂	W₂₄₃	W₂₄₄	W₂₄₅	W₂₄₆	W₂₄₇	W₂₄₈	W₂₄₉	W₂₅₀	W₂₅₁	W₂₅₂	W₂₅₃	W₂₅₄	W₂₅₅	W₂₅₆	W₂₅₇	W₂₅₈	W₂₅₉	W₂₆₀	W₂₆₁	W₂₆₂	W₂₆₃	W₂₆₄	W₂₆₅	W₂₆₆	W₂₆₇	W₂₆₈	W₂₆₉	W₂₇₀	W₂₇₁	W₂₇₂	W₂₇₃	W₂₇₄	W₂₇₅	W₂₇₆	W₂₇₇	W₂₇₈	W₂₇₉	W₂₈₀	W₂₈₁	W₂₈₂	W₂₈₃	W₂₈₄	W₂₈₅	W₂₈₆	W₂₈₇	W₂₈₈	W₂₈₉	W₂₉₀	W₂₉₁	W₂₉₂	W₂₉₃	W₂₉₄	W₂₉₅	W₂₉₆	W₂₉₇	W₂₉₈	W₂₉₉	W₃₀₀	W₃₀₁	W₃₀₂	W₃₀₃	W₃₀₄	W₃₀₅	W₃₀₆	W₃₀₇	W₃₀₈	W₃₀₉	W₃₁₀	W₃₁₁	W₃₁₂	W₃₁₃	W₃₁₄	W₃₁₅	W₃₁₆	W₃₁₇	W₃₁₈	W₃₁₉	W₃₂₀	W₃₂₁	W₃₂₂	W₃₂₃	W₃₂₄	W₃₂₅	W₃₂₆	W₃₂₇	W₃₂₈	W₃₂₉	W₃₃₀	W₃₃₁	W₃₃₂	W₃₃₃	W₃₃₄	W₃₃₅	W₃₃₆	W₃₃₇	W₃₃₈	W₃₃₉	W₃₄₀	W₃₄₁	W₃₄₂	W₃₄₃	W₃₄₄	W₃₄₅	W₃₄₆	W₃₄₇	W₃₄₈	W₃₄₉	W₃₅₀	W₃₅₁	W₃₅₂	W₃₅₃	W₃₅₄	W₃₅₅	W₃₅₆	W₃₅₇	W₃₅₈	W₃₅₉	W₃₆₀	W₃₆₁	W₃₆₂	W₃₆₃	W₃₆₄	W₃₆₅	W₃₆₆	W₃₆₇	W₃₆₈	W₃₆₉	W₃₇₀	W₃₇₁	W₃₇₂	W₃₇₃	W₃₇₄	W₃₇₅	W₃₇₆	W₃₇₇	W₃₇₈	W₃₇₉	W₃₈₀	W₃₈₁	W₃₈₂	W₃₈₃	W₃₈₄	W₃₈₅	W₃₈₆	W₃₈₇	W₃₈₈	W₃₈₉	W₃₉₀	W₃₉₁	W₃₉₂	W₃₉₃	W₃₉₄	W₃₉₅	W₃₉₆	W₃₉₇	W₃₉₈	W₃₉₉	W₄₀₀	W₄₀₁	W₄₀₂	W₄₀₃	W₄₀₄	W₄₀₅	W₄₀₆	W₄₀₇	W₄₀₈	W₄₀₉	W₄₁₀	W₄₁₁	W₄₁₂	W₄₁₃	W₄₁₄	W₄₁₅	W₄₁₆	W₄₁₇	W₄₁₈	W₄₁₉	W₄₂₀	W₄₂₁	W₄₂₂	W₄₂₃	W₄₂₄	W₄₂₅	W₄₂₆	W₄₂₇	W₄₂₈	W₄₂₉	W₄₃₀	W₄₃₁	W₄₃₂	W₄₃₃	W₄₃₄	W₄₃₅	W₄₃₆	W₄₃₇	W₄₃₈	W₄₃₉	W₄₄₀	W₄₄₁	W₄₄₂	W₄₄₃	W₄₄₄	W₄₄₅	W₄₄₆	W₄₄₇	W₄₄₈	W₄₄₉	W₄₅₀	W₄₅₁	W₄₅₂	W₄₅₃	W₄₅₄	W₄₅₅	W₄₅₆	W₄₅₇	W₄₅₈	W₄₅₉	W₄₆₀	W₄₆₁	W₄₆₂	W₄₆₃	W₄₆₄	W₄₆₅	W₄₆₆	W₄₆₇	W₄₆₈	W₄₆₉	W₄₇₀	W₄₇₁	W₄₇₂	W₄₇₃	W₄₇₄	W₄₇₅	W₄₇₆	W₄₇₇	W₄₇₈	W₄₇₉	W₄₈₀	W₄₈₁	W₄₈₂	W₄₈₃	W₄₈₄	W₄₈₅	W₄₈₆	W₄₈₇	W₄₈₈	W₄₈₉	W₄₉₀	W₄₉₁	W₄₉₂	W₄₉₃	W₄₉₄	W₄₉₅	W₄₉₆	W₄₉₇	W₄₉₈	W₄₉₉	W₅₀₀	W₅₀₁	W₅₀₂	W₅₀₃	W₅₀₄	W₅₀₅	W₅₀₆	W₅₀₇	W₅₀₈	W₅₀₉	W₅₁₀	W₅₁₁	W₅₁₂	W₅₁₃	W₅₁₄	W₅₁₅	W₅₁₆	W₅₁₇	W₅₁₈	W₅₁₉	W₅₂₀	W₅₂₁	W₅₂₂	W₅₂₃	W₅₂₄	W₅₂₅	W₅₂₆	W₅₂₇	W₅₂₈	W₅₂₉	W₅₃₀	W₅₃₁	W₅₃₂	W₅₃₃	W₅₃₄	W₅₃₅	W₅₃₆	W₅₃₇	W₅₃₈	W₅₃₉	W₅₄₀	W₅₄₁	W₅₄₂	W₅₄₃	W₅₄₄	W₅₄₅	W₅₄₆	W₅₄₇	W₅₄₈	W₅₄₉	W₅₅₀	W₅₅₁	W₅₅₂	W₅₅₃	W₅₅₄	W₅₅₅	W₅₅₆	W₅₅₇	W₅₅₈	W₅₅₉	W₅₆₀	W₅₆₁	W₅₆₂	W₅₆₃	W₅₆₄	W₅₆₅	W₅₆₆	W₅₆₇	W₅₆₈	W₅₆₉	W₅₇₀	W₅₇₁	W₅₇₂	W₅₇₃	W₅₇₄	W₅₇₅	W₅₇₆	W₅₇₇	W₅₇₈	W₅₇₉	W₅₈₀	W₅₈₁	W₅₈₂	W₅₈₃	W₅₈₄	W₅₈₅	W₅₈₆	W₅₈₇	W₅₈₈	W₅₈₉	W₅₉₀	W₅₉₁	W₅₉₂	W₅₉₃	W₅₉₄	W₅₉₅	W₅₉₆	W₅₉₇	W₅₉₈	W₅₉₉	W₆₀₀	W₆₀₁	W₆₀₂	W₆₀₃	W₆₀₄	W₆₀₅	W₆₀₆	W₆₀₇	W₆₀₈	W₆₀₉	W₆₁₀	W₆₁₁	W₆₁₂	W₆₁₃	W₆₁₄	W₆₁₅	W₆₁₆	W₆₁₇	W₆₁₈	W₆₁₉	W₆₂₀	W₆₂₁	W₆₂₂	W₆₂₃	W₆₂₄	W₆₂₅	W₆₂₆	W₆₂₇	W₆₂₈	W₆₂₉	W₆₃₀	W₆₃₁	W₆₃₂	W₆₃₃	W₆₃₄	W₆₃₅	W₆₃₆	W₆₃₇	W₆₃₈	W₆₃₉	W₆₄₀	W₆₄₁	W₆₄₂	W₆₄₃	W₆₄₄	W₆₄₅	W₆₄₆	W₆₄₇	W₆₄₈	W₆₄₉	W₆₅₀	W₆₅₁	W₆₅₂	W₆₅₃	W₆₅₄	W₆₅₅	W₆₅₆	W₆₅₇	W₆₅₈	W₆₅₉	W₆₆₀	W₆₆₁	W₆₆₂	W₆₆₃	W₆₆₄	W₆₆₅	W₆₆₆	W₆₆₇	W₆₆₈	W₆₆₉	W₆₇₀	W₆₇₁	W₆₇₂	W₆₇₃	W₆₇₄	W₆₇₅	W₆₇₆	W₆₇₇	W₆₇₈	W₆₇₉	W₆₈₀	W₆₈₁	W₆₈₂	W₆₈₃	W₆₈₄	W₆₈₅	W₆₈₆	W₆₈₇	W₆₈₈	W₆₈₉	W₆₉₀	W₆₉₁	W₆₉₂	W₆₉₃	W₆₉₄	W₆₉₅	W₆₉₆	W₆₉₇	W₆₉₈	W₆₉₉	W₇₀₀	W₇₀₁	W₇₀₂	W₇₀₃	W₇₀₄	W₇₀₅	W₇₀₆	W₇₀₇	W₇₀₈	W₇₀₉	W₇₁₀	W₇₁₁	W₇₁₂	W₇₁₃	W₇₁₄	W₇₁₅	W₇₁₆	W₇₁₇	W₇₁₈	W₇₁₉	W₇₂₀	W₇₂₁	W₇₂₂	W₇₂₃	W₇₂₄	W₇₂₅	W₇₂₆	W₇₂₇	W₇₂₈	W₇₂₉	W₇₃₀	W₇₃₁	W₇₃₂	W₇₃₃	W₇₃₄	W₇₃₅	W₇₃₆	W₇₃₇	W₇₃₈	W₇₃₉	W₇₄₀	W₇₄₁	W₇₄₂	W₇₄₃	W₇₄₄	W₇₄₅	W₇₄₆	W₇₄₇	W₇₄₈	W₇₄₉	W₇₅₀	W₇₅₁	W₇₅₂	W₇₅₃	W₇₅₄	W₇₅₅	W₇₅₆	W₇₅₇	W₇₅₈	W₇₅₉	W₇₆₀	W₇₆₁	W₇₆₂	W₇₆₃	W₇₆₄	W₇₆₅	W₇₆₆	W₇₆₇	W₇₆₈	W₇₆₉	W₇₇₀	W₇₇₁	W₇₇₂	W₇₇₃	W₇₇₄	W₇₇₅	W₇₇₆	W₇₇₇	W₇₇₈	W₇₇₉	W₇₈₀	W₇₈₁	W₇₈₂</th

Best-of visual-only analysis

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

- **MediaMill TRECVID2006 baseline run**
- ✓ OK for some concepts, but not spectacular



- All submitted runs should improve upon baseline
- ✓ Three will be highlighted

Proto-concept clustering

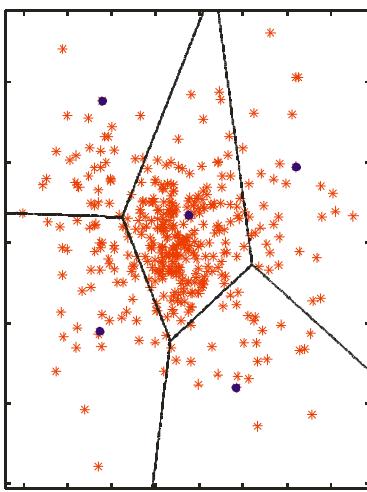
- Introduction
- Challenge
- Visual-only
- Results
- Lessons

► **Problem: pre-defined codebook is restrictive**

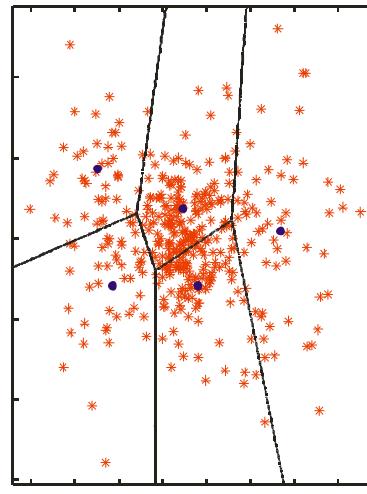
- ✓ Use data-driven codebook instead
- ✓ Tested for regional Weibull and Gabor features only

► **Data-driven prototypes: radius clustering**

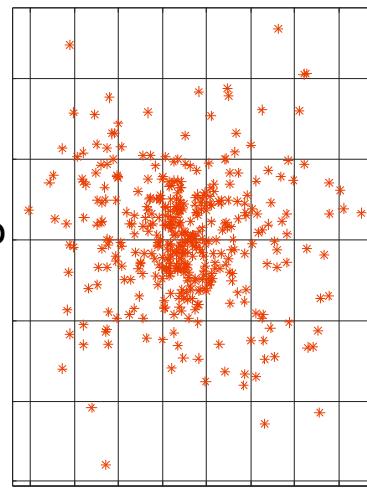
Radius



K-means

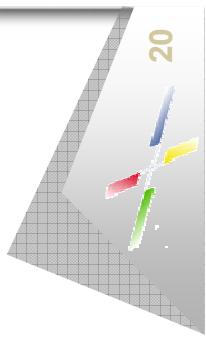


Histogram



► **Investigated in parallel to other experiments**

- ✓ No fusion used

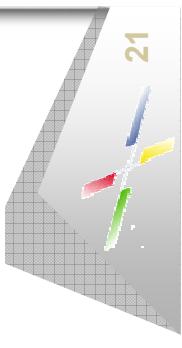
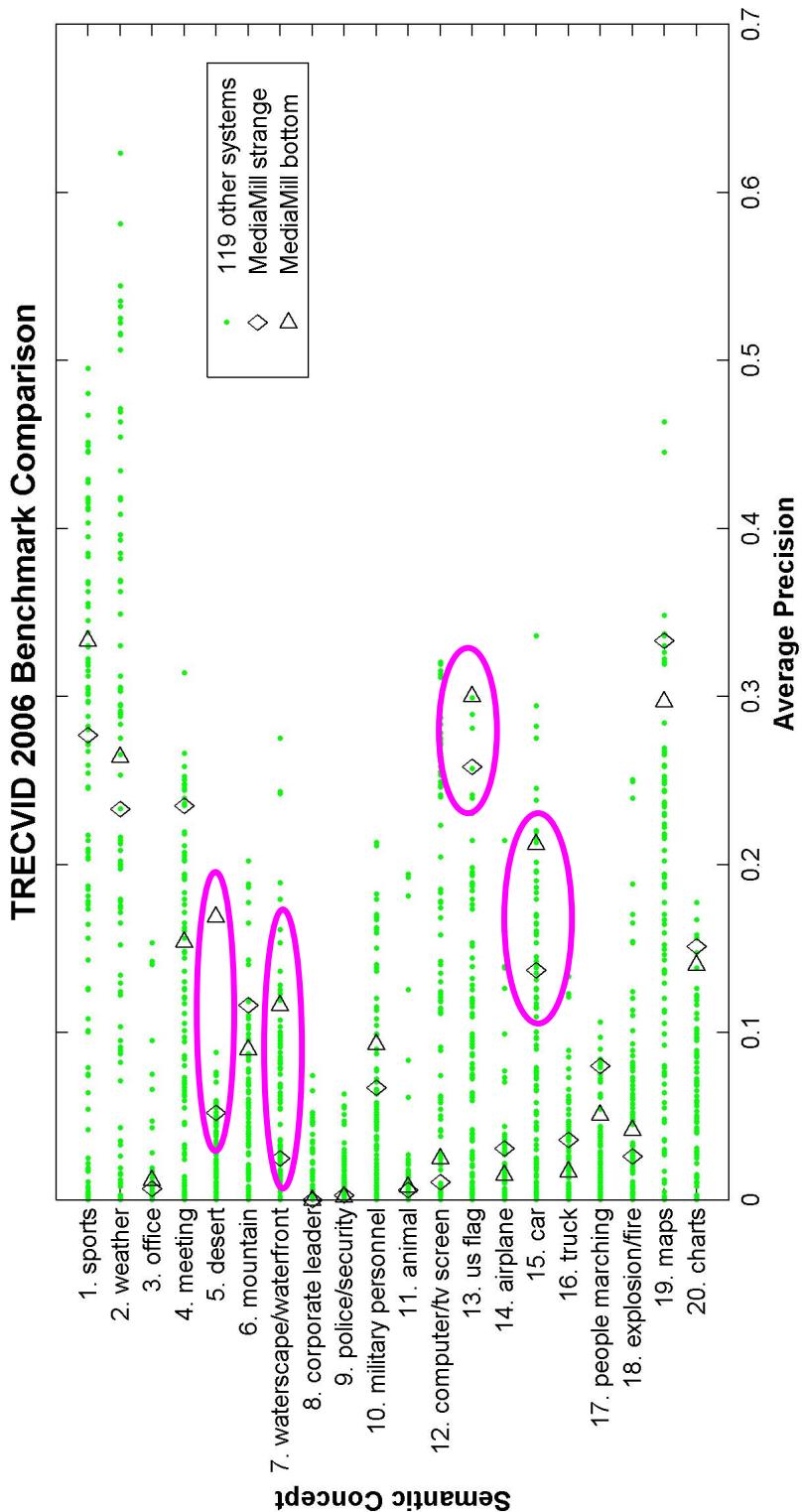


Proto-concept clustering

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

- Outperforms ‘best-of visual’ for half of the concepts
- ✓ Desert = overall best

run: bottom



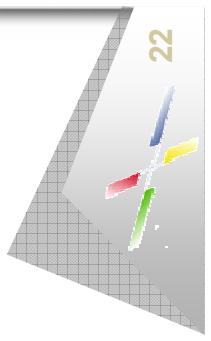
Coloring keypoints

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

➤ Extending SIFT with color pays off

on Challenge

Detector	Descriptor	MAP (Challenge-39)	Change
Harris-Laplace	SIFT	0,198	-
Harris-Laplace	Hue+SIFT	0,231	+16%
Boosted ColorHarris-Laplace	SIFT	0,170	-14%
Boosted ColorHarris-Laplace	Hue	0,169	-15%
Harris-Laplace and Boosted ColorHarris-Laplace	SIFT	0,212	+7%
Late fusion		0,270	+36%

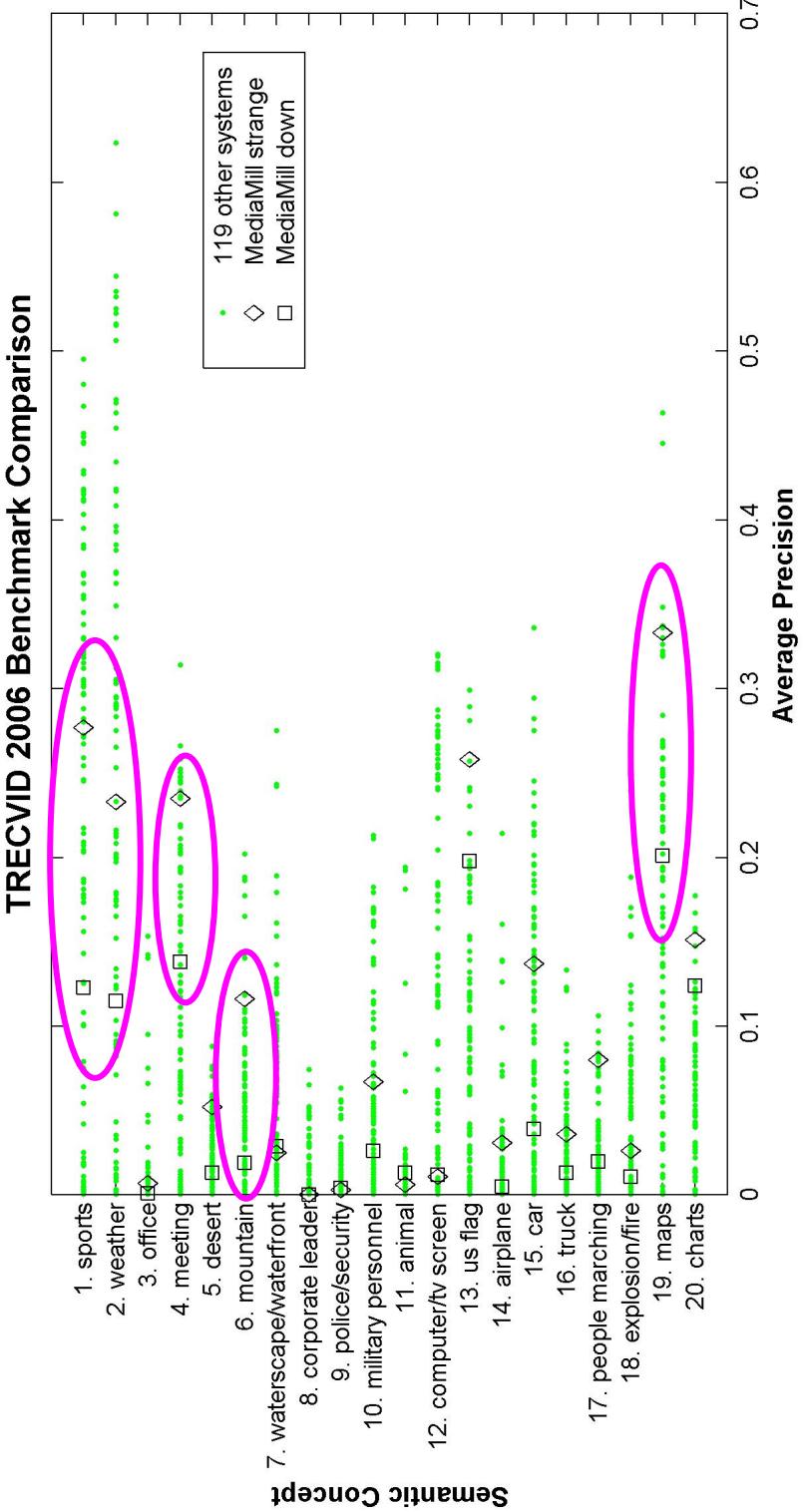


Coloring keypoints

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

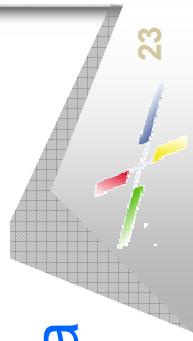
► Modest performance

✓ Unstable for images with only few interest regions



run: down

► A **keypoint-only approach seems a bad idea**

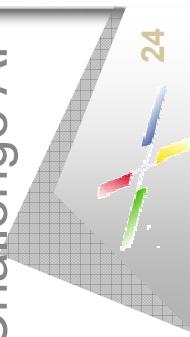


Late fusion of visual-only analysis

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

- All experiments yield for each shot a confidence $p(\omega_j | \vec{x}_i)$
- Fuse concept detection scores using geometric mean
- Lessons
 - ✓ Requires no additional learn set
 - ✓ Favors single high confidence over many low confidences
- Pathfinder Procedure
- Per concept
- Include keypoint run as 1 experiment
- Determine 9 additional experiments based on Challenge AP
- Compute optimal path of experiments based on Challenge AP

$$\exp \left[\frac{1}{n} \sum_{k=1}^n \ln p_k(\omega_j | \vec{x}_i) \right]$$



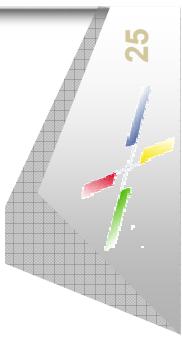
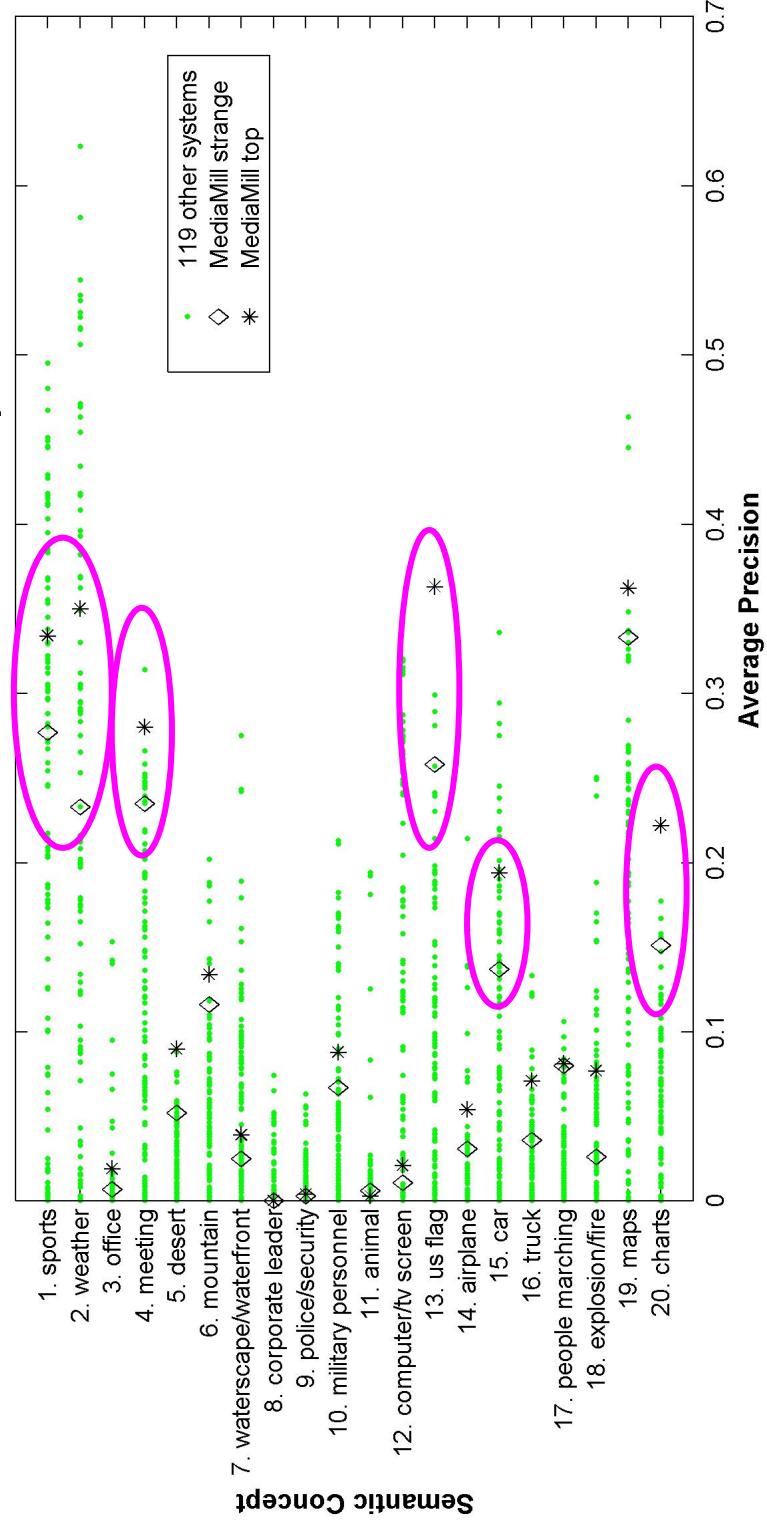
Late fusion of visual-only analysis

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

► **Fusion always improves over best single approach**

✓ Overall best for US flag and Charts

TRECVID 2006 Benchmark Comparison

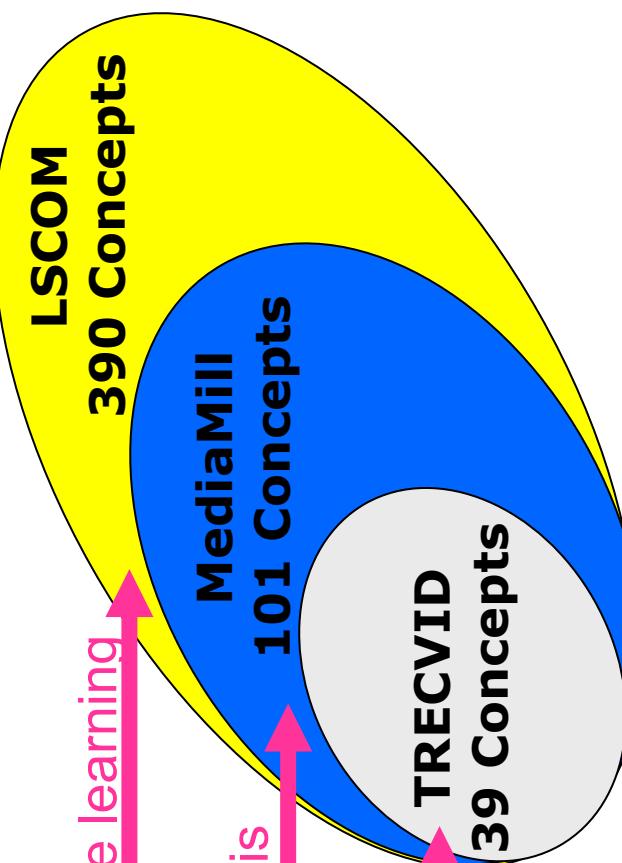


Scaling-up to 491 concept detectors

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

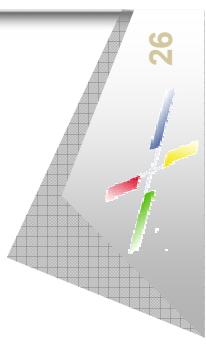
► Graceful degradation

- ✓ Not all experiments were available for all concepts



► Performance not optimal

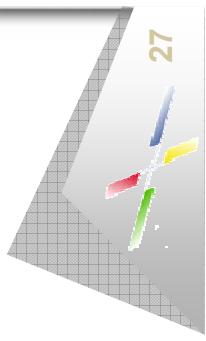
- ✓ Detectors might be useful for semantic video retrieval



Lessons learned

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

- **The MediaMill Challenge allows to**
 - ✓ Gain insight in intermediate video analysis steps
 - ✓ Foster repeatability of experiments
 - ✓ Optimize video analysis systems on a **component level**
 - ✓ Compare and improve upon baseline
- **For visual-only analysis we learned that**
 - ✓ A combination of various techniques pays off
 - ✓ Regional image features seem most effective
 - ✓ Data-driven clustering is more effective than fixed codebook
 - ✓ Keypoint methods unstable for images with few interest points
 - ✓ High-dimensional feature vectors can be handled effectively by relatively simple classifiers like Fisher's linear discriminant
 - ✓ Fusion using geometric mean is cheap and effective
 - ✓ Scaling-up to 1,000+ detectors is a matter of annotated examples



Download location:

www.mediamill.nl/challenge/

- Introduction
- Challenge
- Visual-only
- Results
- Lessons

The screenshot shows the MediaMill Semantic Video Search Engine website. The main navigation bar includes links for "Challenge", "System Demonstration", "Team", and "Publications". Below this, a secondary navigation bar has "Data Set" as the active tab, along with "FAQ" and "Acknowledgment". A sidebar on the left lists "Beständen", "Bereich", "Sage", "Blätterzettel", "Extra", and "Help". The main content area is titled "Systematic Evaluation" and contains a detailed text about the challenge problem for generic video indexing. At the bottom right of the content area, there is a "Klaar" button.

Systematic Evaluation

The purpose of the challenge problem for generic video indexing is to provide researchers with a framework for the systematic evaluation of video indexing components. To allow for systematic evaluation, we organize the challenge problem as a laboratory test. In such a test the variability stemming from multimedia data, concepts, experiments, and performance must be structured to allow for comparison of results. To arrive at a laboratory test for the challenge problem, we separate a multimedia archive in a training set and a test set, using camera shots as the unit for indexing and evaluation, in line with the common procedure in literature. For each set, we provide manually labeled ground truth, at the shot level, in the form of a shared concept lexicon. We define a set of experiments which index shots in the test set based on algorithms tuned on the training set. For each concept in the lexicon this should yield a list of shots, ranked according to detector confidence of concept presence. To evaluate these ranked lists we use average precision.

Multimedia Data Download

The Challenge Problem uses 85 hours of video data from the 2005 NIST TRECVID benchmark (i.e., the TRECVID 2005 training set), containing Arabic, Chinese, and US broadcast news sources, recorded in MPEG-1 during November 2004 by the Linguistic Data Consortium.

Download

If you want to do visual and/or textual feature extraction, and you have not participated in TRECVID 2005 or 2006, the TRECVID data (black hyperlinks) is required.

- **Video data** (Available for fee from LDC, soon)
- **Key frames by Dublin City University** (Available for fee from LDC, soon)
- **Master-Shot-Boundaries by Frammhofer Institute** (Available for fee from NIST)

Check it out and beat the baseline!

